



Improving runoff estimates from regional climate models: a performance analysis in Spain

D. González-Zeas¹, L. Garrote¹, A. Iglesias², and A. Sordo-Ward¹

¹Department of Hydraulic and Energy Engineering, Technical University of Madrid, Spain

²Department of Agricultural Economics and Social Sciences, Technical University of Madrid, Spain

Correspondence to: D. González-Zeas (dunia.gonzalez.zeas@alumnos.upm.es)

Received: 14 November 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 5 January 2012

Revised: 25 May 2012 – Accepted: 29 May 2012 – Published: 25 June 2012

Abstract. An important step to assess water availability is to have monthly time series representative of the current situation. In this context, a simple methodology is presented for application in large-scale studies in regions where a properly calibrated hydrologic model is not available, using the output variables simulated by regional climate models (RCMs) of the European project PRUDENCE under current climate conditions (period 1961–1990). The methodology compares different interpolation methods and alternatives to generate annual times series that minimise the bias with respect to observed values. The objective is to identify the best alternative to obtain bias-corrected, monthly runoff time series from the output of RCM simulations. This study uses information from 338 basins in Spain that cover the entire mainland territory and whose observed values of natural runoff have been estimated by the distributed hydrological model SIMPA. Four interpolation methods for downscaling runoff to the basin scale from 10 RCMs are compared with emphasis on the ability of each method to reproduce the observed behaviour of this variable. The alternatives consider the use of the direct runoff of the RCMs and the mean annual runoff calculated using five functional forms of the aridity index, defined as the ratio between potential evapotranspiration and precipitation. In addition, the comparison with respect to the global runoff reference of the UNH/GRDC dataset is evaluated, as a contrast of the “best estimator” of current runoff on a large scale. Results show that the bias is minimised using the direct original interpolation method and the best alternative for bias correction of the monthly direct runoff time series of RCMs is the UNH/GRDC dataset, although the formula proposed by Schreiber (1904) also gives good results.

1 Introduction

In recent decades, many studies have addressed the potential impact of climate change on hydrology and water resources (i.e. Arnell, 1999; Vicuna and Dracup, 2007; Buytaert et al., 2009; Elsner et al., 2010), gaining considerable attention among researchers. The use of regional climate models (RCMs) is an important tool for assessing water management under future climate change scenarios (Varis et al., 2004). In order to evaluate this impact in hydraulic systems, it is necessary to have monthly time series of stream flow for the period of analysis. The standard procedure to study climate change impacts on water resources is to downscale climatic variables, such as temperature or precipitation, from RCMs, and use a hydrologic rainfall-runoff model to generate stream flow series for the basin of interest. However, it is very difficult to calibrate a rainfall runoff model for regional studies covering extensive areas. If a calibrated rainfall-runoff model is not available, the only source of stream flow time series may be other indirect variables simulated by the RCMs, such as surface runoff, whose generation process is conceptually similar to that of hydrologic rainfall-runoff models. These variables may reproduce the basic seasonal and spatial characteristics of surface hydrology (Graham et al., 2007; Hurk et al., 2004) and could be useful to undertake water availability analyses. In addition to that, the UNH/GRDC (University of New Hampshire/Global Runoff Data Center) gives high-resolution annual and monthly mean surface runoff datasets that preserve the accuracy of measurements of observed flows at the main hydrologic stations around the world. Currently, this layer is the “best estimator”

of surface runoff over great extensions of land (Fekete et al., 1999), and it can be used to correct bias from RCMs.

In light of this situation, it is essential to analyse several alternatives that will enable us to obtain simulated stream flow series and can be introduced in water resources models to estimate climate change impacts. In the present study, we consider the use of direct surface runoff series simulated using RCMs or obtained applying climate formulas (such as those based on the aridity index; Arora, 2002), which give surface runoff variables on an annual time scale. In this regard, it is noteworthy that numerous studies have highlighted the calculation of the mean annual surface runoff using a non-parametric focus that does not depend on a hydrologic model (i.e. Sankarasubramanian and Vogel, 2003), using climate formulas based on the aridity index to evaluate the sensitivity of surface runoff to climate changes (in 1337 basins in the United States).

In hydrologic applications, the spatial resolution of a given climate model plays a significant role in determining its impact on hydrologic systems. For this reason, in recent years, the use of RCMs has improved the spatial details of climate change projections obtained with global climate models (Déqué et al., 2005). The European PRUDENCE project (Christensen et al., 2007) provides high-resolution climate change scenarios with results that include the simulations of several RCMs for the projection of climate variables in current and future climate situations. However, despite the relatively high resolution of RCMs, hydrologic modelling generally requires information on a smaller scale than that provided by the typical RCM grid size (Hagemann and Gates, 2003). Thus, one of the limiting factors regarding the use of information projected by RCMs (in hydrologic predictions) is the scale misalignment between the output of the climate models and the scale used by the hydrologic models (the river basin) (e.g. Lettenmaier et al., 1999; Bergström et al., 2001; Wood et al., 2002). Several studies have designed different downscaling techniques to overcome this deficiency, focusing their attention on temperature and precipitation variables because they are the key variables by which hydrologic models calculate surface runoff (e.g. Leung et al., 2003; Dibike and Coulibaly, 2005; Terink et al., 2010).

A broad range of downscaling techniques has been developed, including methods involving artificial neural networks (Hewitson and Crane, 1992), empirical regression methods (Von Storch et al., 1993) and methods based on linear interpolation, spatial disaggregation and bias correction (Wood et al., 2004; Chen et al., 2011). For all of these, the general limitations are well documented in the existing literature (Wilby et al., 2004). However, many of these reports only evaluate the advantages and disadvantages associated with a specific use of the method and do not give accurate information regarding the best method for evaluating the effects of climate change on hydrologic systems (Fowler et al., 2007). Therefore, using several methodological alternatives to downscale runoff output from RCMs in current climate conditions and

comparing their results with respect to observed data may reveal the procedure that gives the best fit.

Another factor that limits the use of RCM output is its bias, which has to be eliminated before it can be used for other purposes (Christensen et al., 2008). In general, the behaviour of climate models shows less bias for precipitation and temperature than for surface runoff (Girorgi et al., 1994; González-Zeas, 2010). Bias-correction techniques have been developed for precipitation and temperature variables used in hydrologic models (Piani et al., 2010a, b). Changes in specific statistical aspects, generally mean and variance values, of the simulated fields are used in the bias correction formulations, which are directly applied to current observations that are subsequently used to force the hydrologic models (Haerter et al., 2011). Many hydrologic models have been used to determine current surface runoff values from precipitation and temperature variables simulated by climate models (e.g. Leavesley et al., 2002; Döll et al., 2003). However, due to the limited availability of data and the lack of a correctly calibrated hydrologic model that allows surface runoff determination from basic climate variables (i.e. precipitation and temperature) when working on a large regional or continental scale, it is often difficult to generate surface runoff series that accurately represent the current conditions and, thus, can be successfully employed with water resources management models (Strzepek and Yates, 1997; Yates, 1997; Hagemann et al., 2004; Kirchner, 2006; Silberstein, 2006; Döll et al., 2008). For this reason it is important to evaluate which of the alternatives considered in this study minimises bias with respect to observed values.

Our study focuses on providing estimates for runoff conditions in current situation using the RCM outputs. Nonetheless, if the RCM runoff is to be used in climate change impact analysis, the first step is to analyze how the RCM simulations reproduce the current situation, and therefore our methodology focuses on the control scenario as a first step before linking to climate change impact analysis.

The present study compares several interpolation procedures used to downscale the surface runoff variable by applying them to projections made via 10 RCM simulations from the PRUDENCE European project under current climate conditions. The capability of each method to reproduce the observed behaviour of this variable is emphasised. Conducted for 338 basins that cover the whole of mainland Spain, the study's main objective is to determine the accuracy of surface runoff output obtained from RCM simulations and to evaluate possible alternatives for generating monthly series that minimise bias with respect to the observed values. The specific objectives are as follows: (1) to determine the interpolation method that most closely approximates the observed values, (2) to evaluate to what extent the outputs of direct surface runoff from RCMs and climate formulas can be used for large-scale studies in the absence of correctly calibrated hydrologic models, (3) to validate the methods by comparing results to global runoff data provided by the UNH/GRDC and

(4) to generate a monthly series that is representative of current conditions using a simple bias correction methodology based on the alternative that best reproduces the observed values.

2 Data

2.1 Area of study

The area of study is the mainland territory of Spain, which has an area of 504 782 km². The study considers 338 sub-basins, which are defined from points in the river network that are relevant to the management of water resources (MARM, 1998). Figure 1 shows the 338 elemental sub-basins. The basins on which this study has been performed were obtained by accumulating all the elemental sub-basins located upstream of the point being considered. Table 1 summarises the characteristics of the accumulated basins and classifies them by river basin districts (RBDs).

The basins studied are broadly variable in their size, with the largest being 84 923.85 km² (the basin of the Ebro RBD) and the smallest being 15.89 km² (in the North I RBD). The mean size of the basins is 4358.13 km², with a coefficient of variation of 1.89.

2.2 Observed surface runoff series

The hydrologic regimes of most rivers in Spain are strongly affected by water abstractions and flow regulations. Thus, data measured in the gauging stations cannot be directly compared to surface runoff in natural regimes. Natural regime surface runoff data were obtained from the results of the Precipitation Runoff Integrated Model (SIMPA, by its Spanish acronym) (Estrela and Quintas, 1996). This model is calibrated over 100 control points of the entire territory of Spain, using stations where stream flows are measured in natural regimes (MARM, 2000; Alvarez et al., 2004), enabling the monthly series of runoff in natural regimes to be obtained in a grid with a 1 km² resolution. Evaluation/comparisons of the results of the climate models employed the monthly surface runoff series generated by the SIMPA model (labelled as observed data in this study) in each of the 338 elemental study basins for the period from 1961–1990.

2.3 Regional climate models

The present study uses projections made by eight regional climate models presented in the European PRUDENCE project (Christensen et al., 2007) nested in a single global model, referred to as HadAM3H. The regional models give information relative to a large number of variables (temperature, precipitation, runoff, evaporation and solar radiation, among others) with daily, monthly and seasonal resolution for the periods from 1961–1990 (control period) and 2071–2100 (climate change scenario). The PRUDENCE project

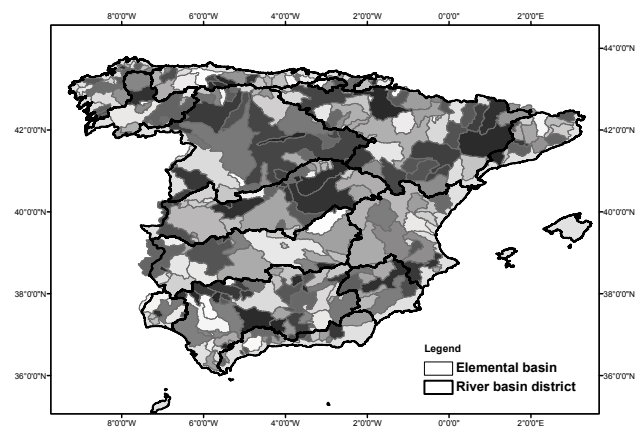


Fig. 1. Elemental basins of study. The methods used in this study were applied to the accumulated basins below, including all the sub-basins upstream of the point being considered.

gives the values of the simulated climate variables, using the grid of the original coordinates (O) of each model, with a spatial resolution that varies between 25–50 km depending on the model, and in the unified coordinates CRU (Climate Research Unit) with a spatial resolution of $0.5 \times 0.5^\circ$ (cells of approximately 50 km sides). The DMI model has 3 different simulations; thus, in total, this study has used data generated by 10 climate simulations of the PRUDENCE project (Table 2).

3 Methodology

The methodology used herein consists of comparing the observed surface runoff series with the series downscaled from the results of the RCMs via different alternatives and analysing the results to select the most accurate procedure. Given that the SIMPA model has been specifically conceived and calibrated to generate monthly runoff series with high spatial resolution, its results should be more accurate than those produced by RCMs. The surface runoff layers generated by the UNH/GRDC, an additional global reference, are also considered because they specifically pertain to surface runoff with the same spatial resolution as generated by RCMs.

The runoff series in the basins were generated from the results of the RCMs of the PRUDENCE project using the following two procedures: (1) directly using the monthly runoff variables of the models and (2) using five climate formulas based on the aridity index. The average values of the basic variables in each basin (surface runoff, temperature, precipitation and solar radiation) were obtained using a spatial interpolation procedure in a grid with greater resolution than the simulations of the models, considering the outputs of the variables in original coordinates (O) of each model as well as the unified CRU coordinates. The following calculation

Table 1. Characteristics of the basins used in this study by river basin district.

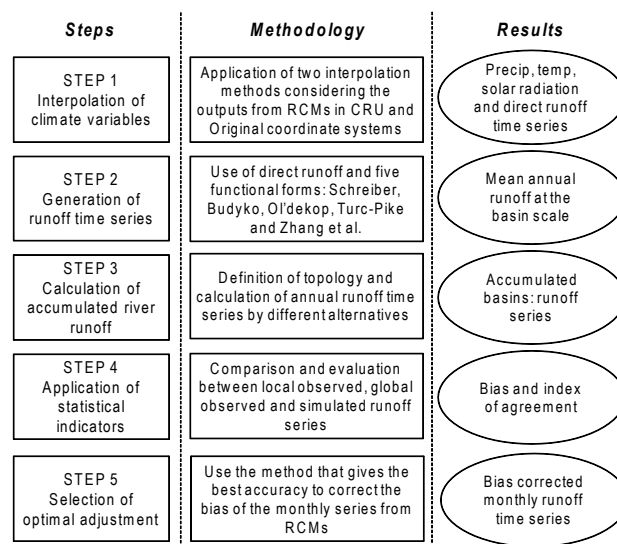
| River Basin District | Code | Number of basins | Maximum area (km ²) | Minimum area (km ²) | Mean area (km ²) | Coefficient of variation |
|----------------------|--------|------------------|---------------------------------|---------------------------------|------------------------------|--------------------------|
| North I | NorI | 15 | 16 320.85 | 15.89 | 2295.24 | 1.86 |
| North II | NorII | 20 | 13 888.75 | 31.36 | 1468.41 | 2.04 |
| North III | NorIII | 13 | 4463.18 | 31.31 | 737.24 | 1.57 |
| Galician Coast | Gal | 20 | 8599.34 | 31.09 | 786.69 | 2.38 |
| Ebro | Ebr | 48 | 84 923.85 | 31.42 | 6186.81 | 2.56 |
| Catalunya | Cat | 16 | 12 458.64 | 32.32 | 1931.93 | 1.53 |
| South | Sou | 20 | 9973.39 | 68.70 | 980.90 | 2.22 |
| Segura | Seg | 9 | 15 752.37 | 738.34 | 5494.19 | 1.09 |
| Jucar | Juc | 17 | 35 912.64 | 50.25 | 7586.91 | 1.28 |
| Duero | Due | 33 | 75 544.62 | 63.08 | 8693.35 | 2.37 |
| Tagus | Taj | 39 | 55 758.17 | 16.25 | 6587.23 | 2.18 |
| Guadiana I | GnaI | 23 | 54 395.58 | 99.70 | 9806.86 | 1.68 |
| Guadiana II | GnaII | 9 | 6303.90 | 101.81 | 1253.73 | 1.61 |
| Guadalquivir | Gqv | 56 | 60 939.80 | 34.19 | 7204.32 | 2.11 |
| Total | | 338 | 84 923.85 | 15.89 | 4358.13 | 1.89 |

procedures were used in each case: (1) direct method (D) and (2) interpolated method (I). Topological analyses were performed on the elemental basins, and surface runoff values were obtained for the accumulated basins. The goodness of fit between the observed and generated annual runoff time series was evaluated using statistical indicators. Next, we determined and evaluated the alternative that best reproduced the observed values and that was, therefore, used to correct the bias and to generate the monthly surface runoff series at the hydrologic basin scale (which can be introduced in the water resources management models). As summarised in Fig. 2, the analysis process had the following steps: (1) interpolation of climate variables, (2) generation of surface runoff series for the basins studied, (3) generation of surface runoff series for the accumulated basins, (4) determination of the goodness of fit of the generated surface runoff series and (5) determination of the method and alternative that provided the best agreement with observed values and, thus, that was used to correct the bias of the monthly series simulated by the RCMs.

3.1 Interpolation methods

Many interpolation methods have been developed to span the gap from climate models to basin scale. Some of these statistical approaches used a truncated Gaussian weighting filter (Thornton et al., 1997), PRISM method (Daly et al., 2002), the thin-plate smoothing splines procedure (Hijmans et al., 2005), among others.

We have used four procedures to generate the basic series for the basins studied. The base was determined by the outputs of the 10 RCM simulations in the O coordinates of each model and in the common CRU coordinate system with a resolution of $0.5 \times 0.5^\circ$. The behaviours of the two coordinate systems were evaluated, because, although it is easier to

**Fig. 2.** Scheme of the analysis methodology.

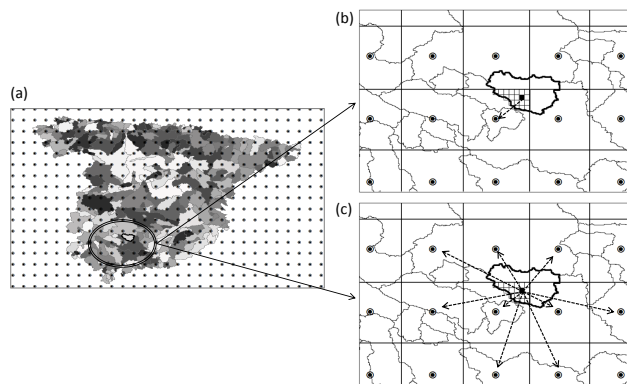
work in the common system, the interpolation process that enables transition to the unified CRU domain can alter the values of some variables, particularly in the coastal regions. In addition, errors introduced by the existence of different landscapes at specific points between the original grid and the regular 0.5° grid can be overlooked (Hagemann and Jacob, 2007).

In general, the spatial resolution of the cells of the RCMs is inadequate when working at the river basin scale. Given that the area of some sub-basins can be less than the area of the output grids of the RCMs, spatial disaggregation of the cells is necessary to achieve a finer scale that can take into account the smallest sub-basins. A 2.5 min size grid was used for the

Table 2. Summary of the regional climate model simulations used in this study and produced by the European PRUDENCE Project (Christensen et al., 2007).

| Driving Global Climate Model | *Centre or Institute | Regional Climate Models (RCMs) | Simulations Control Scenario (1961–1990) | Acronym |
|------------------------------|----------------------|--------------------------------|--|---------|
| HadAM3H | DMI | HIRHAM | HC1 | DMI.1 |
| | | | HC2 | DMI.2 |
| | | | HC3 | DMI.3 |
| | ETH | CHRM | HC_CTL | ETH |
| | | | CTL | GKSS |
| | | | ref | ICTP |
| | KNMI | RACMO | HC1 | KNMI |
| | | | 3003 | MPI |
| | | | HCCTL | SMHI |
| | UCM | PROMES | control | UCM |

*DMI: Danish Meteorological Institute; ETH: Swiss Federal Institute of Technology; GKSS: Institute of Coastal Research; ICTP: Physics of Weather and Climate Section; KNMI: The Royal Netherlands Meteorological Institute; MPI: Max-Planck-Institute for Meteorology; SMHI: Swedish Meteorological and Hydrological Institute; UCM: Universidad Complutense de Madrid.

**Fig. 3.** Interpolation methods: (a) CRU domain of the RCMs, (b) direct method (D) and (c) interpolated method (I). A similar procedure was used with the RCM outputs in original coordinates.

working scale. The outputs of the RCMs were translated into this grid with two interpolation methods: the D method and the I method. Figure 3 shows the scheme used; Fig. 3a shows the CRU grid superimposed on the elemental sub-basins of the analysis; Fig. 3b shows the application of the D method, in which the assumed value of the variable on the work grid is the nearest element of the grid RCM; Fig. 3c shows the application of the I method, in which the assumed value on the work grid is obtained from the values for the nine nearest points in the RCM grid through a weighted mean, in which the weighting coefficient of each value is the inverse of the distance squared. A similar procedure is used with the output in O coordinates. Therefore, four interpolation methods are compared in this study: CRU-D, CRU-I, O-D and O-I.

3.2 Generated temporal surface runoff series

Using the four interpolation methods, mean annual surface runoff series were generated for the basins studied, considering the following alternatives: (1) the direct runoff of the 10 RCM simulations and (2) the mean annual surface runoff obtained via application of five functional forms based on the aridity index (Arora, 2002). In addition, the D and I procedures were applied to the surface runoff datasets of the UNH/GRDC, a reference of the current global surface runoff (Fekete et al., 1999).

Arora (2002) showed that the use of functional forms based on the aridity index (ϕ), defined as the ratio between the potential evapotranspiration (PET) and precipitation (P), is a reasonable first-order approximation of the actual evapotranspiration (ET). Functional forms based on this index consist of formulas proposed by different authors (Schreiber, 1904; Ol'dekop, 1911; Budyko, 1948; Turc, 1954-Pike, 1964; Zhang et al., 2001; Table 3) that allow calculation of the mean annual surface runoff values. These formulas determine the ratio between the actual evapotranspiration (ET) and the precipitation (P) through the balance of water and energy.

The mean value of annual surface runoff (R) is obtained by the balance of water as follows:

$$R = P - ET - \Delta S. \quad (1)$$

Assuming that ΔS (the change in moisture storage in soil) is very small over an annual time scale, the actual evapotranspiration (ET) can be calculated in each of the functional forms of Table 3 and substituted in Eq. (1), thereby yielding the value of the mean annual surface runoff as a function of the aridity index:

$$R = P(1 - F(\phi)), \quad (2)$$

Table 3. Functional forms $F(\phi)$ used to calculate the evapotranspiration ratio.

| Functional Form | Evapotranspiration ratio |
|-------------------------|--|
| Schreiber (1904) | $\frac{ET}{P} = 1 - e^{-\phi}$ |
| Ol'dekop (1911) | $\frac{ET}{P} = \phi \tanh\left(\frac{1}{\phi}\right)$ |
| Budyko (1948) | $\frac{ET}{P} = \left[\phi \tanh\left(\frac{1}{\phi}\right) \left(1 - e^{-\phi}\right)\right]^{1/2}$ |
| Turc (1954)-Pike (1964) | $\frac{ET}{P} = \frac{1}{\sqrt{0.9 + \left(\frac{1}{\phi}\right)^2}}$ |
| Zhang et al. (2001) | $\frac{ET}{P} = \frac{1 + w\phi}{1 + w\phi + \frac{1}{\phi}}$ w : coefficient that depends on the kind of vegetation |

where $F(\phi)$ corresponds to each of the five functional forms used.

For calculation of the PET, the Hargreaves method is used (Hargreaves and Samani, 1982) as follows:

$$PET = 0.0023 \cdot (T_{\text{mean}} + 17.8) \cdot (T_{\text{max}} - T_{\text{min}})^{0.5} R_A, \quad (3)$$

where PET is the potential evapotranspiration in mm, R_A is the solar radiation in the upper part of the atmosphere in W m^{-2} , T_{mean} is the mean temperature in $^{\circ}\text{C}$, T_{max} is the maximum temperature in $^{\circ}\text{C}$ and T_{min} is the minimum temperature in $^{\circ}\text{C}$.

Information regarding precipitation, temperature and solar radiation used for the calculation of the PET and the mean annual surface runoff is obtained from the outputs of the PRUDENCE RCM simulations. As a consequence, for each functional form of Table 3, 10 mean annual surface runoff values are generated, 1 for each simulation of the PRUDENCE project.

The UNH/GRDC dataset provides monthly climatological runoff fields, which are runoff outputs from a water balance model that is driven by observed meteorological data and then corrected with the runoff fields that are disaggregated from the observed river discharges at the main gauging stations in the world. The UNH/GRDC dataset preserves the accuracy of the observed discharge measurements and maintains the spatial and temporal distributions of the simulated surface runoff. Thus, because it provides the “best estimate” of surface runoff on a large scale (Fekete et al., 2002), the UNH/GRDC dataset has been used as a reference in the calibration and validation of different surface runoff models (e.g. SIMTOP, SIMGM) that use global and regional climate models (Wei et al., 2002; Niu et al., 2005; Sperna Weiland et al., 2012; Wisser et al., 2010). This study used the annual surface runoff layer of the UNH/GRDC dataset as a reference for the “best results” that can be obtained at the working scale of each RCM. The surface runoff layer of the UNH/GRDC dataset has the same resolution as the CRU grid.

A total of 60 mean annual surface runoff values for each interpolation method are generated as follows: 10 corresponding to the RCM simulations of the European PRUDENCE project and 50 corresponding to the 5 climate formulas (10 for each functional form). In addition, two annual runoff series (D and I) are generated using the UNH/GRDC runoff fields. In total, 40 series of direct runoff are obtained, considering that four methods of interpolation are used. Similarly, 200 series are obtained for the five climatic formulas and 2 series for the UNH/GRDC dataset, since in this case only a single simulation is available and the outputs are only in CRU coordinate systems.

3.3 Criteria for comparison(s) of results

It is appropriate to evaluate/compare models, such as those used in this study, using quantitative indicators of goodness of fit (Legates and McCabe, 1999). Along these lines, analyses of the discrepancies between the output of the models and the observed values have been reported by several authors (Adam and Lettenmaier, 2003; Fekete and Vörösmarty, 2004). Thus, quantitative measurements were used to determine the agreement (goodness of fit) between the mean annual surface runoff observed and the simulated surface runoff determined by different interpolation methods, and the bias and the index of agreement proposed by Willmott (1981) were evaluated.

The bias reveals the model's tendency to overestimate or underestimate one variable and quantifies the systematic error of the model (Janssen and Heuberger, 1995). The bias can be determined from the mean error, which is normalised by the mean of the values observed for the group of 338 accumulated basins and for the different alternatives analysed as follows:

$$\text{Bias} = \frac{\bar{S} - \bar{O}}{\bar{O}}, \quad (4)$$

where \bar{S} and \bar{O} represent the mean annual simulated and observed runoff for the group of basins studied, respectively.

In order to obtain a better measure of the relative agreement between simulations and observations, the index of agreement was also determined. The index of agreement is a measure of the mean relative error obtained by normalising the mean quadratic error with respect to the potential error, which represents the largest value that the squared difference of each pair can attain (sum of the absolute value of the difference between the predictions and the mean of the observations and the observations and the mean of the observations). As a dimensionless measure that can be used to compare models and that is broadly discussed in Willmott (1982), Willmott et al. (1985), Hall (2001), Krause et al. (2005) and Moriasi et al. (2007), the index of agreement is expressed in the following manner:

$$d = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (|S_i - \bar{O}| + |O_i - \bar{O}|)^2}, \quad (5)$$

where S_i represents the simulated surface runoff in basin i , O_i is the surface runoff observed in the basin i , \bar{O} is the mean observed surface runoff, n is the number of basins studied and d is the index of agreement (which varies between 0, not a good adjustment, and 1, perfect adjustment).

3.4 Correction of the bias

Considering the lack of a correctly calibrated hydrologic model in large-scale studies, monthly surface runoff series are built from RCM simulations, using the alternative that best reproduces the observed values. The results are validated with information obtained from the UNH/GRDC dataset. It is well known that the output of the RCMs cannot be used directly if there is no procedure that eliminates the existing bias (Sharma et al., 2007). Studies conducted by Murphy (1999), Kidson and Thompson (1998) and Wilby et al. (2000) suggest the need to correct the bias in the outputs of the climate models to ensure accurate results in hydrology and water resources management applications. Different methodologies have been used to correct the bias. The simplest methodology uses changes in one specific statistical aspect, often the mean or the variance. This method is equivalent to correcting the observed series with a constant multiplying factor (e.g. Klein et al., 2005). More advanced bias-correcting methodologies, such as those developed by Schmidli et al. (2006) and Leander and Buishand (2007), correct the bias using more than one specific statistical aspect. Improvements in the simpler methods of bias correction have been obtained by fitting the probability density function to the histograms of the observed data (Piani et al., 2010a; Maraun et al., 2010), the application of the quantile mapping method (Wilby et al., 2000; Déqué, 2007; Bardossy and Pegram, 2011). This study proposes a simple bias-correction methodology based on the determination of an annual correction factor, which is obtained using the alternative that minimises the bias and gives the surface runoff series that most closely agrees with observed values. Using the mean of the annual series as a representative statistical parameter, constant multiplying factors are determined that are then used to correct the monthly runoff time series directly simulated via RCMs. The corresponding corrected monthly runoff series are obtained as follows:

$$R_{\text{control.cor},i} = R_{\text{control},i} \cdot \left(\frac{R_{\text{alt}}}{R_{\text{control}}} \right), \quad (6)$$

where $R_{\text{control.cor},i}$ are the corrected monthly runoff series of the RCMs, $R_{\text{control},i}$ are the monthly runoff series of the RCMs, R_{alt} is the mean annual runoff of the alternative that

best approximates the observed values and R_{control} is the mean annual runoff value of the RCMs.

In order to evaluate the performance of the results obtained for the corrected monthly surface runoff series, the Kolmogorov-Smirnov (K-S) non-parametric test (Conover, 1980) is used to compare monthly cumulative distribution function (cdf) with the observed runoff.

The K-S test can be described as follows. Suppose $F_1(x)$ and $F_2(x)$ are two cdfs of two sample data of a variable x . The null hypothesis and the alternative hypothesis concerning their cdfs are

$$H_0 : F_1(x) = F_2(x) \text{ for all } x$$

$H_A : F_1(x) \neq F_2(x)$ for at least one value of x and the test statistics, T is defined as

$$T = \sup_x |F_1(x) - F_2(x)|, \quad (7)$$

which is the maximum vertical distance between the distributions $F_1(x)$ and $F_2(x)$. If the test statistics is greater than the critical value, the null hypothesis is rejected.

3.5 Regime curves

To represent the annual hydrological cycle once the annual bias of the series has been corrected, we calculated the hydrological regime curve which consists of the 30-yr average mean monthly runoff, obtained for all 12 months individually from observed and simulated runoff in Spain. Regime curves have been calculated from observed runoff, direct runoff obtained by each RCM and bias corrected runoff by UNH/GRDC and Schreiber (1904) formula for the 10 RCM simulations. Additionally to plot the regime curves of Spain, the Nash-Sutcliffe (NS) coefficient of efficiency is used to compare the 30-yr average monthly values displayed in the regime curves (e.g. Wolock and McCabe, 1999; Sperna Weiland et al., 2010). Besides, we have taken into account the performance spatially, calculating the NS in the 14 river basin districts of Spain.

NS is a dimensionless indicator widely used to evaluate the goodness of fit of the data (Nash and Sutcliffe, 1970). This coefficient represents the relative improvement of the simulated values with respect to the mean of the observed values and is expressed in the following manner:

$$NS = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}, \quad (8)$$

where S_i and O_i represent the simulated and observed values in month i , and \bar{O} is the mean observed surface runoff. NS ranges from minus infinity (poor model) to 1 (perfect model).

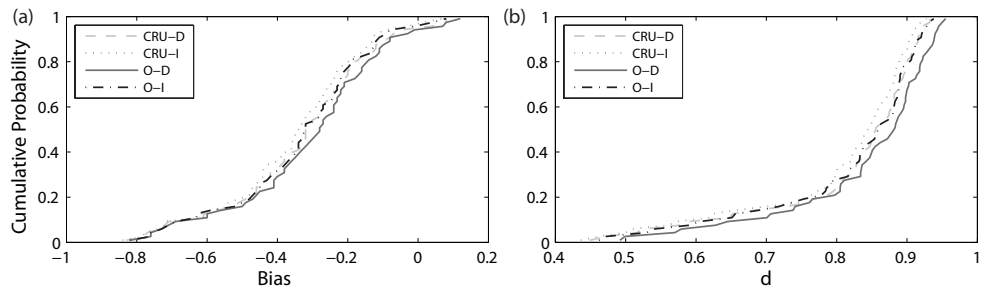


Fig. 4. Cumulative probability of the 60 values that correspond to all of the alternatives analysed for each interpolation method: **(a)** bias and **(b)** index of agreement.

Table 4. Mean values of the bias and the index of agreement of the annual mean surface runoff obtained with the 10 RCMs for each interpolation method and for the different alternatives considered. Bold numbers are used to remark the best results.

| Alternatives | Bias (%) | | | | Index of Agreement (d) | | | |
|-------------------------|-----------|-------|------------|-----|----------------------------|-------|--------------|-------|
| | CRU-D | CRU-I | O-D | O-I | CRU-D | CRU-I | O-D | O-I |
| Direct Runoff | −36 | −38 | −35 | −37 | 0.761 | 0.746 | 0.779 | 0.762 |
| Schreiber (1904) | −23 | −25 | −20 | −23 | 0.867 | 0.855 | 0.883 | 0.867 |
| Ol’dekop (1911) | −47 | −49 | −44 | −47 | 0.775 | 0.756 | 0.804 | 0.775 |
| Budyko (1948) | −34 | −37 | −32 | −35 | 0.830 | 0.815 | 0.852 | 0.830 |
| Turc (1954)–Pike (1964) | −36 | −38 | −33 | −36 | 0.823 | 0.808 | 0.846 | 0.823 |
| Zhang et al. (2001) | −25 | −27 | −22 | −25 | 0.857 | 0.845 | 0.874 | 0.857 |
| Mean | −33 | −36 | −31 | −34 | 0.819 | 0.804 | 0.840 | 0.819 |
| UNH/GRDC | −3 | −4 | | | 0.959 | 0.957 | | |

4 Results

4.1 Comparison of the interpolation methods

In this section, the interpolation method that generates a surface runoff series that most closely approximates the observed values is determined. Using the bias and the index of agreement, the relative errors of the mean annual runoff obtained by the different alternatives were calculated for each interpolation procedure. The values of the indicators represent the behaviour of the group of the 338 basins of the study.

Figure 4 shows the cumulative probability of the bias and index of agreement of the CRU-D, CRU-I, O-D and O-I methods for 10 simulations of direct surface runoff and 5 climate formulas. In total, 60 values are presented for each interpolation method. The cumulative probability of the bias curve (Fig. 4a) shows that the bias of the O-D method is less than those of all other methods, indicating that the O-D method tends to minimise errors with respect to observed values. Similarly, the results shown in Fig. 4b indicate that the cumulative probability of d , the index of agreement, obtained with the O-D method is greater than that obtained by any other method, indicating a better adjustment.

Table 4 summarises the values of the indicators for the mean of the 10 RCM obtained for each alternative using each

interpolation method. Once again, the best results are obtained with the O-D method, which outputs mean bias and d -values of -31% and 0.840 , respectively. However, the difference between the O-D method and other methods is not very marked; comparing the O-D method to the CRU-D, O-I and CRU-I methods, the error in the results decreases by only 2% , 3% and 5% , respectively.

Similar comparisons were made with the mean annual surface runoff values given by the UNH/GRDC dataset. For methods D and I, the results show an existing bias with respect to the observed values of -3% and -4% and d -values of 0.959 and 0.957 , respectively, indicating that method D gives the best results.

4.2 Evaluation of the annual surface runoff obtained by direct runoff from RCMs and climate formulas

In this section, the direct runoff values determined by the RCMs and the mean annual runoff obtained from five climate formulas are compared to determine if they can adequately represent the observed values of mean annual surface runoff. The results are also compared to the reference surface runoff values obtained from the UNH/GRDC dataset.

We have already determined that the O-D method is the method that minimises bias with respect to the observed

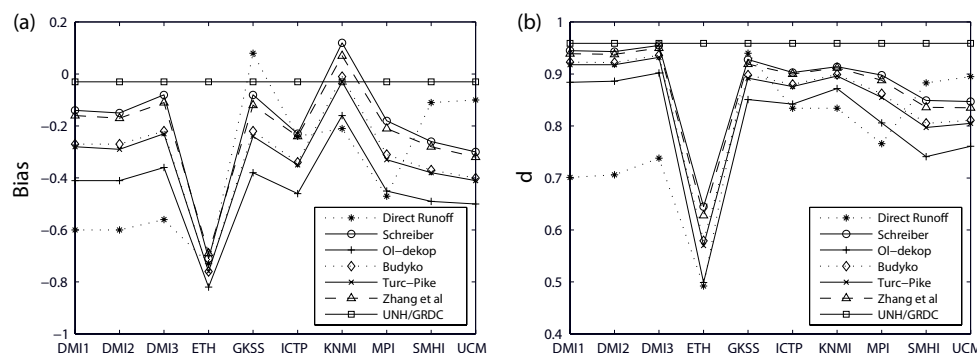


Fig. 5. Goodness of fit indicators obtained for the direct surface runoff, the five functional forms (Schreiber, 1904; Ol'dekop, 1911; Budyko, 1948; Turc, 1954-Pike, 1964; Zhang et al., 2001) for the 10 RCM simulations and for the UNH/GRDC dataset: (a) bias and (b) t -index of agreement. The acronyms of the RCM simulations are defined in Table 2.

values. Therefore, from this point on, we use the mean annual surface runoff results calculated with this procedure. Figure 5a and b show the fitting indicators of the direct surface runoff from RCMs, the runoff from the climate formulas of Schreiber (1904), Ol'dekop (1911), Budyko (1948), Turc (1954)-Pike (1964), Zhang et al. (2001) and the mean annual runoff in the UNH/GRDC dataset. Results are shown for all 10 RCM simulations. The fitting indicators show best behaviour for the UNH/GRDC dataset, because the associated bias is minimum with a value of -3% and the d -value is 0.959, which is greater than the d -values obtained with alternative approaches. The climate formulas behave better with regard to the direct runoff of the RCMs. Compared to the rest of the functional forms, Schreiber's formula performs better with regard to all RCMs simulations studied. The formula from Zhang et al. (2001) shows similar results, although they are slightly less accurate than the ones obtained with Schreiber's formula. Using the formulas of Budyko (1948) and Turc (1954)-Pike (1964), similar results are generated, and results using Ol'dekop's formula show the poorest agreement of all the functional forms. Despite demonstrating the least favourable overall results, comparisons with direct surface runoff demonstrate better agreements than Schreiber's formula for the GKSS, SMHI and UCM models. The bias obtained for the direct surface runoff is -35% for the mean of the 10 RCMs, while that obtained for Schreiber's formula is -20% .

Some of the 10 RCMs obviously provide better results than others. With regard to Schreiber's formula, the best performing model is DMI.3, which has a bias of -8% and a d -value of 0.955. Regarding direct surface runoff, the GKSS model performs best with bias and d -values of 8% and 0.940, respectively. With all alternatives investigated, the ETH model had the greatest errors when compared to the observed surface runoff in the basins of Spain.

Therefore, Schreiber's formula and the DMI.3 model produce the best fit with observed values. Compared to the UNH/GRDC dataset, the error in the results is on the order of -5% .

4.3 Corrected monthly series

Considering that the monthly series determined via direct runoff from RCMs cannot be directly introduced into the management models of water resources due to their observed bias and that Schreiber's formula gives only annual values of surface runoff in the basins studied, this section generates corrected monthly time series from simulations obtained using RCMs. These series are generated using annual correction factors calculated based on the methodology that most accurately approximates the observed values. The results are compared with the monthly time series that are corrected with the annual values given by the global runoff reference of the UNH/GRDC dataset.

So as to compare the corrected monthly distributions with respect to observed and direct runoff, Fig. 6 illustrates the cumulative probability of the monthly runoff obtained by ordering from lowest to highest value of every month of the entire period of analysis (1961–1990). The results in the figure highlight the overall behaviour of the 10 RCM simulations for the entire territory of mainland Spain. The direct runoff values obtained by the 10 RCM simulations in Fig. 6a are significantly lower than the observed data, whereas the corrected values by UNH/GRDC dataset in Fig. 6b and corrected values by Schreiber's formula in Fig. 6c are much closer. In order to evaluate the performance of the probability distribution of the monthly runoff obtained by the direct and bias-corrected values, Table 5 shows the results of non-parametric test at 95 % significance level for the 10 RCM simulations and for the average of them. The K-S test demonstrates that neither the 10 RCM simulations of the direct runoff nor the average of them can reproduce the probability distribution of runoff because all p -values are below 0.05. On the other hand, the corrected runoff improves the performance in the great majority of the RCM simulations; however, only the DMI.2, GKSS and MPI RCMs and the average of the 10 RCMs are able to simulate the probability distribution of the monthly observed runoff.

Table 5. Test results (p-values) of the K-S test for the difference between observed and (a) direct runoff, (b) runoff bias corrected by UNH/GRDC and (c) runoff bias corrected by Schreiber's formula in Spain at 95 % confidence level, for the 10 RCM simulations. Bold characters are used to remark the passed test. The acronyms of the RCM simulations are defined in Table 2.

| Alternatives | DMI1 | DMI2 | DMI3 | ETH | GKSS | ICTP | KNMI | MPI | SMHI | UCM | Mean |
|------------------|-----------------------|-----------------------|-----------------------|-----------------------|-------------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|
| Direct Runoff | 1.9×10^{-26} | 1.5×10^{-27} | 6.0×10^{-21} | 8.1×10^{-48} | 0.04 | 1.0×10^{-4} | 2.3×10^{-3} | 1.0×10^{-14} | 3.4×10^{-11} | 9.0×10^{-9} | 1.4×10^{-7} |
| UNH/GRDC | 0.03 | 0.07 | 0.01 | 5.6×10^{-9} | 0.48 | 3.1×10^{-3} | 3.7×10^{-5} | 0.10 | 3.4×10^{-11} | 3.4×10^{-9} | 0.12 |
| Schreiber (1904) | 1.8×10^{-3} | 0.02 | 1.8×10^{-3} | 2.3×10^{-8} | 0.20 | 2.3×10^{-3} | 5.3×10^{-4} | 0.10 | 2.0×10^{-11} | 9.0×10^{-9} | 0.06 |

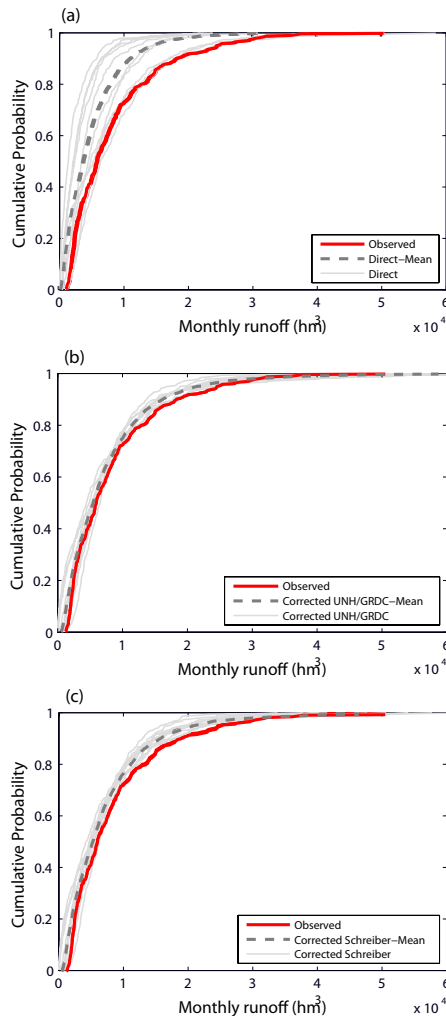


Fig. 6. Monthly distribution of runoff in Spain. The figures show the cumulative probability of the mean runoff of every month of the entire period of analysis (1961–1990). The comparison is made between the observed data and (a) direct runoff, (b) bias corrected runoff by UNH/GRDC dataset and (c) bias corrected runoff by Schreiber (1904). Red line is observed monthly distribution of runoff; grey lines are monthly distributions of runoff simulated by the 10 RCMs, and dotted line is the mean of the simulations.

4.4 Regime curves

Figure 7 shows the mean monthly surface runoff values in Spain for each of the 10 RCM simulations utilised in this study. The observed runoff values are compared to runoff corrected by Schreiber's formula for the DMI.3 RCM simulation, the runoff corrected by UNH/GRDC dataset and direct runoff from RCMs. The monthly values in the figure represent the 30-yr mean behaviour (1961–1990). Generally, the regime curves of the corrected RCM series show a reduction of the differences between the regimes derived from the direct runoff and adequately represent the predominant seasonal cycle of the observed regime. Nonetheless, there are certain seasonal discrepancies that vary from one model to another. The GKSS, SMHI and UCM models more adequately represent the observed seasonal cycle. The DMI.1, DMI.2, DMI.3, KNMI and MPI models tend to underestimate the observed values during autumn and winter, and the ICTP and ETH models underestimate the values in winter and spring.

Figure 8 summarises the NS efficiency coefficients obtained as a result of the comparison between the regime curves (direct runoff and bias corrected runoff with UNH/GRDC dataset and Schreiber's formula) with respect to observed regime curve for the 14 RBDs of Spain and for all of the mainland territory (plot in Fig. 7). The efficiency indicator values show that the results significantly improve with the corrected series. The NS values obtained by the 10 RCM simulations are shown by the boxplot. The negative values of NS indicate the minimal capacity of the simulations to reproduce the observed series. In all the RBDs, direct surface runoff is characterised by presenting the most unfavourable NS values. A clear improvement in the NS values is observed for the corrected series; nonetheless, the efficiency in the results varies spatially and depends on the correction method used. In the case of mainland Spain and the North I, Duero, Tagus, Guadiana I and Guadalquivir RBDs, the series corrected with the UNH/GRDC dataset provide the best results. On the other hand, the series corrected by Schreiber's formula give better results in the North II, Galician Coast, Ebro, internal basins of Catalunya, Jucar and Segura RBDs. The results show negative NS values for both the direct series and the corrected series in the North II, South and Guadiana II RBDs.

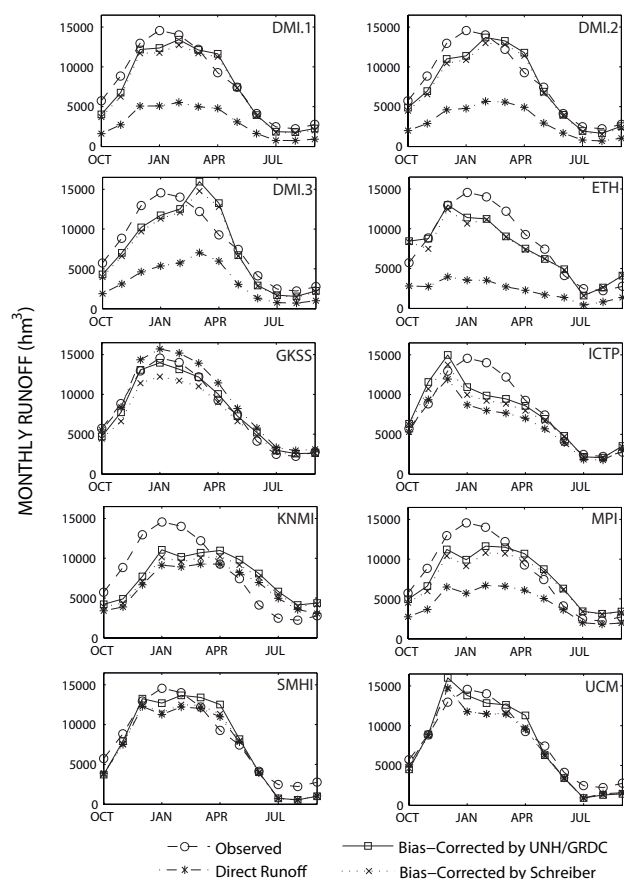


Fig. 7. Comparison of mean monthly runoff in mainland Spain for the 10 RCM simulations obtained from observed values, direct runoff, bias corrected by UNH/GRDC dataset and bias corrected by Schreiber's formula (mean of period 1961–1990). The acronyms of the RCM simulations are defined in Table 2.

5 Discussion and conclusions

This study explores different alternatives that utilise RCM simulations to generate surface runoff time series that approximate observed values and that comprise the basis for generating a design series that, when introduced into management models, will allow us to evaluate the impacts on water availability under future scenarios of climate change. Considering the absence of a correctly calibrated hydrologic model when working on a large scale, this study analysed four interpolation procedures for downscaling surface runoff at the river basin scale. Different alternatives were also considered, particularly with regard to the measure to which each of them approximates the observed values.

Using fit capability indicators, the behaviour of the group of Spanish basins was evaluated using four interpolation procedures: CRU-D, CRU-I, O-D and O-I. The cumulative probability curves (for the bias and the index of agreement values) show that the O-D method minimises deviation with respect to the observed values. This result indicates that the O-D

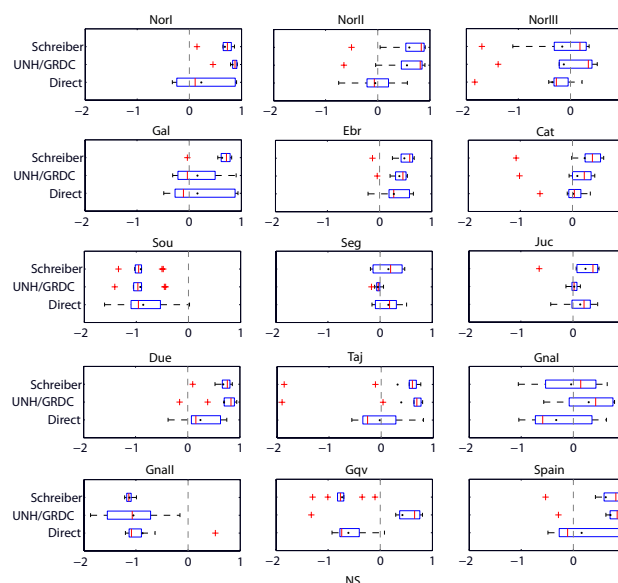


Fig. 8. Nash-Sutcliffe (NS) efficiency coefficients of the 30-yr average monthly mean of direct and corrected surface runoff for Spain and its 14 river basin districts. The boxplots show the NS obtained by the 10 RCM simulations. The lines extend up to 1.5 times the interquartile range to the right and left of the box. The box extends from the 25th percentile to the 75th percentile. The line within the box indicates the median of the simulations. The black point shows the mean of the simulations, and red crosses outside the box indicate the outliers. The river basin district code is defined in Table 1.

method is the best procedure to determine the runoff series for a given river basin using RCM outputs. It is easiest to determine the values of the variables in each basin based on the influence of the nearest point in the grid that corresponds to the climate model simulation. Similar results were obtained for the runoff series generated by the climate formulas and the UNH/GRDC dataset. Others studies (for example, Wood et al., 2004) have also used spatial disaggregation techniques on RCM outputs.

Once the O-D method was established as the best spatial disaggregation process at the river basin scale, different alternatives were compared to determine which of them gives the best approximation to the observed values of mean annual runoff. These alternative methods were evaluated by comparing the direct runoff simulated by the RCMs and the mean annual runoff calculated with climate formulas. The results were contrasted with the surface runoff given by the UNH/GRDC dataset. The statistical indicators applied show that climate formulas based on the aridity index approximate better the observed values than direct runoff obtained from RCM simulations. It is because climate formulas use precipitation and temperature variables to determine mean annual surface runoff values and because the uncertainties associated with climate-model simulations present less bias for precipitation and temperature than for

surface runoff. Other studies (e.g. Sankarasubramanian and Vogel, 2002; Arora, 2002; Potter and Zhang, 2009; McMahon et al., 2011) have also satisfactorily evaluated surface runoff using the formulas proposed by Schreiber (1904), Ol'dekop (1911), Budyko (1948), Turc (1954)-Pike (1964) and Zhang et al. (2001).

Of the five functional forms of the aridity index, Schreiber's formula generates mean annual surface runoff values that minimise error compared to observed data. The least favourable results were obtained with the direct surface runoff method, with the exception of the GKSS, SMHI and UCM models, whose goodness-of-fit indicators show superior behaviour compared to results obtained with Schreiber's formula. In addition, the UNH/GRDC values show better agreement with observed values compared to results obtained with climate formulas and direct runoff. This result confirms that, as established by Fekete et al. (2002), the combination of a water balance model and contributions of the main hydrologic stations of the world can generate mean annual runoff fields that are consistent with observed values, preserving the spatial accuracy for large-scale domains. Comparison of climate formulas and direct runoff reveals that, despite the fact that they were not based on a model specifically conceived for the study of surface runoff, Schreiber's formula applied to the DMI.3 model presents results similar to those obtained with the UNH/GRDC dataset. Similarly, Vogel et al. (1999) and Sankarasubramanian and Vogel (2003) determined that, on a large scale and without requiring the assumption of a model or a calibration strategy, the use of non-parametric estimators, such as those based on the aridity index to determine surface runoff, give results that are comparable to those obtained with a basin model.

The findings reported in this study show that the corrected runoff improves the direct runoff simulated by the RCMs. However, the K-S test of the corrected monthly distribution of runoff indicates that only some of the RCMs and the average of them are capable to reproduce the observed behaviour.

The application of the bias-correction method to the direct runoff of the RCMs decreases the deviation to the 30-yr average observed values. Evaluation of the regime curves of direct runoff simulated by the RCMs and corrected with the UNH/GRDC dataset and Schreiber's formula applied to the DMI.3 model in Spain has demonstrated a greater accuracy in matching the observations using the annual correction factor based on UNH/GRDC dataset. Thus, for the control scenario, the best alternative for correcting the monthly runoff series simulated by RCMs is to use the global surface runoff information given by the UNH/GRDC dataset. However, it is also worth mentioning that the series corrected with Schreiber's formula give good values for the NS coefficient and do not differ significantly from the UNH/GRDC results obtained for mainland Spain. In some cases, the former alternative actually shows improved performance when the behaviour is evaluated at the RBD level. It is noteworthy that, when working with accumulated basins, some errors of

upstream basins can be compensated for downstream basins. The NS efficiency coefficients in Fig. 8 show different behaviours for the different RBDs. Thus, the RBDs that belong to the Cantabrian, Pyrenees and Mediterranean slopes (North II, Galicia, Ebro, Internal Basins of Catalunya, Segura and Jucar) show more accurate results when the bias is corrected with Schreiber's formula. However, the RBDs that belong to the Atlantic slope (Duero, Tagus, Guadiana and Guadalquivir) show more satisfactory results when the bias is corrected with the UNH/GRDC dataset. The different capability of performance obtained by the two alternatives may be caused by the following: (1) the deficiency of the RCMs to simulate the climatic variables in arid and semi-arid regions, (2) the UNH/GRDC dataset has been calibrated in the Spanish gauging stations that belong to the RBDs that give the better performance results with this alternative, (3) the inability of the climate formulas to capture the impact of rapid precipitation events over arid and semi-arid regions and (4) the basin's size. However, both the UNH/GRDC dataset and Schreiber's formula improve the performance of the RCMs and can be used to correct the bias of the direct surface runoff simulations. The existing literature contains several studies that focus on bias correction of the temperature and precipitation variables, the results of which are then used in the hydrologic models to determine surface runoff values (e.g. Fujihara et al., 2008; Piani et al., 2010b). However, this study underscores the importance of correcting the bias of the direct surface runoff simulated by RCMs, suggesting that the accuracy of the results of the corrected monthly series may be comparable to those obtained by other studies (e.g. Hay et al., 2002; Hay and Clark, 2003).

A shortcoming of annual models, such as the formulas based on the aridity index described here, is the inability to track changes in soil moisture and the loss of important seasonal variability which exists in many regions. Because the corrected monthly surface runoff series are obtained from direct runoff in RCM simulations, the monthly, seasonal and year-to-year variability in the current climate situation will depend on the characteristics of each RCM. The results depicted in Fig. 7 show a general trend toward accurate representation of the observed seasonal cycles. However, the results differ from one climate model to another. The GKSS model best represents the observed seasonal cycles for mainland Spain.

Fekete et al. (1999) showed that working with a spatial resolution of $0.5 \times 0.5^\circ$ provides more accurate results in basins with areas greater than 25 000 km². In fact, RBDs with smaller areas, like South and Guadiana II, give less accurate results. Thus, it is important to consider the effects of the study area on the accuracy of the results. Although this particular observation was not an intended objective of this study, it will be kept in mind for future research. In addition, the results from Hagemann and Jacob (2007) indicate that the accuracy of RCM results is different in different basins.

Despite the fact that RCM-generation of surface runoff uses simple simulation methods and is not necessarily designed to calculate the stream flow values with the accuracy of a hydrologic model, results obtained by this method reflect the general trends of water balance. In addition, the errors that can exist can be directly linked to errors within the variables related to the hydrologic cycle, which are errors that cannot be eliminated even by the most sophisticated hydrologic models.

In general, the findings of this study conclude that, for using direct surface runoff outputs from RCMs, the O-D interpolation method at basin scale and the bias-correction with annual factors given by the UNH/GRDC dataset or Schreiber's climate formula alternatives are necessary in order to obtain simulated monthly runoff that most closely approximates the observed values.

Acknowledgements. This research has been supported by MAEC – AECID (Ministry of Foreign Affairs and Cooperation – Spanish Agency for International Cooperation) through its scholarships for foreigners to doctoral studies, European Commission CIRCE project (contract no. 036961), ARCO project (200800050084350) of the Spanish Ministry of Environment, Rural and Marine Affairs (MARM) and CYTED Action VIAGUA (410AC0399). Data have been provided through the PRUDENCE data archive, funded by the European Union through contract EVK2-CT2001-00132.

Edited by: M. Vanclooster

References

- Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, *J. Geophys. Res.*, 108, 42–57, 2003.
- Álvarez, J., Sánchez, A., and Quinta, L.: SIMPA, a GRASS based tool for Hydrological Studies, Proceedings of the FOSS/GRASS users Conference, Bangkok, Thailand, 12–14 September, 2004.
- Arnell, N. W.: Climate change and water resources, *Global Environ. Change*, 9, S31–S49, 1999.
- Arora, V. K.: The use of the aridity index to assess climate change effect on annual runoff, *J. Hydrol.* 265, 164–177, 2002.
- Bárdossy, A. and Pegram, G.: Downscaling precipitation using regional climate models and circulation patterns towards hydrology, *Water Resour. Res.*, 47, W04505, doi:10.1029/2010WR009689, 2011.
- Bergström, S., Carlsson, B., Gardelin, M., Lindström, G., Pettersson, A., and Rummukainen, M.: Climate change impacts on runoff in Sweden-assessments by global climate models, dynamical downscaling and hydrological modeling, *Clim. Res.*, 16, 101–112, 2001.
- Budyko, M. I.: Evaporation under natural conditions, *Gidrometeorizdat*, Leningrad, USSR, English translation by IPST, Jerusalem, 1948.
- Buytaert, W., Célleri, R., and Timbe, L.: Predicting climate change impacts on water resources in the tropical Andes: the effects of GCM uncertainty, *Geophys. Res. Lett.*, 36, L07406, doi:10.1029/2008GL037048, 2009.
- Chen, J., Brissette, F. P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, *J. Hydrol.*, 401, 190–202, 2011.
- Christensen, J. H., Carter, T. R., Rummukainen, M., and Amanatidis, G.: Evaluating the performance and utility of regional climate models: the PRUDENCE Project, *Climate Change*, 81, 1–6, 2007.
- Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophys. Res. Lett.*, 35, L20709, doi:10.1029/2008GL035694, 2008.
- Conover, W. J.: Practical Nonparametric Statistics, 2nd Edn. Wiley, New York, 1980.
- Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., and Pasteris, P.: A knowledge-based approach to the statistical mapping of climate, *Clim. Res.*, 22, 99–113, 2002.
- Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, *Global Planet. Change*, 57, 16–26, 2007.
- Déqué, M., Jones, R. G., Wild, M., Giorgi, F., Christensen, J. H., Hassell, D. C., Vidale, P. L., Rockel, B., Jacob, D., Kjellström, E., Castro, M., Kucharski, F., and van den Hurk, B.: Global high resolution versus limited area model climate change projections over Europe: quantifying confidence level from PRUDENCE results, *Clim. Dynam.*, 25, 653–670, 2005.
- Dibike, Y. B. and Coulibaly, P.: Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models, *J. Hydrol.*, 307, 145–163, 2005.
- Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, *J. Hydrol.*, 270, 105–134, 2003.
- Döll, P., Berkhoff, K., Bormann, H., Fohrer, N., Gerten, D., Hagemann, S., and Krol, M.: Advances and visions in large-scale hydrological modelling: findings from the 11th Workshop on Large-Scale Hydrological Modelling, *Adv. Geosci.*, 18, 51–61, doi:10.5194/adgeo-18-51-2008, 2008.
- Elsner, M. M., Cuo, L., Voisin, N., Deems, J. S., Hamlet, A. F., Vano, J. A., Mickelson, K. E., Lee, S., and Lettenmaier, D. P.: Implications of 21st century climate change for the hydrology of Washington State, *Climate Change*, 102, 225–260, 2010.
- Estrela, T. and Quintas, L.: El sistema integrado de modelización precipitación aportación (SIMPA), *Revista de Ingeniería Civil*, 104, 43–52, 1996.
- Fekete, B. M. and Vörösmarty, C. J.: Uncertainties in precipitation and their impacts on runoff estimates, *J. Climate*, 17, 294–304, 2004.
- Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: Global, Composite Runoff Fields Based on Observed River Discharge and Simulated Water Balances, Rep. 22, Global Runoff data Cent., Koblenz, Germany, available at: <http://www.grdc.sr.unh.edu/html/paper/ReportUS.pdf> (last access: 20 June 2011), 1999.
- Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining river discharge and simulated water balances, *Global Biogeochem. Cy.*, 16, 1042, doi:10.1029/1999GB001254, 2002.
- Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recently advances in downscaling techniques for hydrological modeling, *Int. J. Climatol.*, 2002.

- 27, 1547–1578, 2007.
- Fujihara, Y., Tanaka, K., Watanabe, T., Nagano, T., and Kojiri, T.: Assessing the impacts of climate change of the Seyhan River Basin in Turkey: Use of dynamically downscaled data for hydrologic simulations, *J. Hydrol.*, 353, 33–48, 2008.
- Graham, L. P., Hageman, S., Jaun, S., and Beniston, M.: On interpreting hydrological change from regional climate models, *Climate Change*, 81, 97–122, 2007.
- Giorgi, F., Hostetler, S. W., and Brodeur, C. S.: Analysis of the surface runoff hydrology in a regional climate model, *Quart. J. Roy. Meteor. Soc.*, 120, 161–183, 1994.
- González-Zeas, D.: Análisis hidrológico de los escenarios de cambio climático en España. Unpublished DEA from Ph.D. Thesis, Technical University of Madrid, Spain, 2010.
- Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, *Hydrol. Earth Syst. Sci.*, 15, 1065–1079, doi:10.5194/hess-15-1065-2011, 2011.
- Hagemann, S. and Gates, D.: Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations, *Clim. Dynam.*, 21, 349–359, 2003.
- Hagemann, S. and Jacob, D.: Gradient in the climate change signal of European discharge predicted by a multi-model ensemble, *Climate Change*, 81, 309–327, 2007.
- Hagemann, S., Machehauer, B., Jones, R., Christensen, O. B., Déqué, M., Jacob, D., and Vidale, P. L.: Evaluation of water and energy budgets in regional climate models applied over Europe, *Clim. Dynam.*, 23, 547–567, 2004.
- Hall, M. J.: How well does you model fit the data?, *J. Hydroinform.*, 3, 49–55, 2001.
- Hargreaves, G. H. and Samani, Z. A.: Estimating potential evapotranspiration, *J. Irrig. Drain. E.-ASCE*, 108, 225–230, 1982.
- Hay, L. E. and Clark, M. P.: Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States, *J. Hydrol.*, 282, 56–75, 2003.
- Hay, L. E., Clark, M. P., Wilby, R. L., Gutowski Jr., W. J., Leavesley, G. H., Pan, Z., Arritt, R. W., and Takle, E. S.: Use of regional climate model output for hydrologic simulations, *J. Hydrometeorol.*, 3, 571–590, 2002.
- Hewitson, B. C. and Crane, R. G.: Large-scale atmospheric controls on local precipitation in tropical Mexico, *Geophys. Res. Lett.*, 19, 1835–1838, 1992.
- Hijmas, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, *Int. J. Climatol.*, 25, 1965–1978, 2005.
- Hurk, B., Hirschi, M., Schär, C., Lenderink, G., Meijgaard, E., Ulden, A., Rockel, B., Hagemann, S., Gahan, P., Kjellström, E., and Jones, R.: Soil control on runoff response to climate change in regional climate model simulations, *J. Climate*, 18, 3536–3551, 2004.
- Janssen, P. H. M. and Heuberger, P. S. C.: Calibration of process-oriented models, *Ecol. Model.*, 83, 55–66, 1995.
- Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, *Adv. Geosci.*, 5, 89–97, doi:10.5194/adgeo-5-89-2005, 2005.
- Kidson, J. W. and Thompson, C. S.: A comparison of statistical and model-based downscaling techniques for estimating local climate variations, *J. Climate*, 11, 735–753, 1998.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements analysis and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362, 2006.
- Klein, J., Frei, C., Gurtz, J., Lüthi, D., Vidale, P. L., and Schär, C.: Hydrologic simulations in the Rhine basin driven by a regional climate model, *J. Geophys. Res.*, 110, D04102, doi:10.1029/2004JD005143, 2005.
- Leander, R. and Buishand, T. A.: Re-sampling of regional climate model output for the simulation of extreme river flows, *J. Hydrol.*, 332, 487–496, 2007.
- Leavesley, G. H., Markstrom, S. L., Restrepo, P. J., and Viger, R. J.: A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling, *Hydrol. Process.*, 16, 173–187, 2002.
- Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of “goodness of fit” measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35, 233–241, 1999.
- Leung, L. R., Mearns, L. O., Giorgi, F., and Wilby, R. L.: Regional climate research, *B. Amer. Meteorol. Soc.*, 84, 89–95, 2003.
- Lettenmaier, D. P., Wood, A. W., Palmer, R. N., Wood, E. F., and Stakhiv, E. Z.: Water resources implications of global warming: A U.S. regional perspective, *Climatic Change*, 43, 537–579, 1999.
- MacMahon, T., Peel, M. C., Pegram, G. G. G., and Smith, I. N.: A simple methodology for estimating mean and variability of annual runoff and reservoir yield under present and future climates, *J. Hydrometeorol.*, 12, 135–146, 2011.
- MARM: Metodologías de Evaluación de los Sistemas de Explotación de Recursos Hídricos, Centro de Estudios y Experimentación de Obras Públicas (CEDEX), Ministerio de Medio Ambiente y Medio Rural y Marino, Spain, 1998.
- MARM, Ministry of Environment, Ministerio de Medio Ambiente: White Paper Book of Water in Spain. Libro Blanco del Agua en España. Secretaría de Estado de Aguas y Costas, available at: <http://www.mma.es/rechid/librob/index.htm> (last access: 10 July 2011), 2000.
- Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themessl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, 48, RG3003, doi:10.1029/2009RG000314, 2010.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Velth, T.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *T. ASABE*, 50, 885–900, 2007.
- Murphy, J.: An evaluation of statistical and dynamical techniques for downscaling local climate, *J. Climate*, 12, 2256–2284, 1999.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Niu, G. Y., Yang, Z. L., Dickinson, R. E., and Gulden, L. E.: A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, *J. Geophys. Res.*, 110, D21106, doi:10.1029/2005JD006111, 2005.

- Ol'dekop, E. M.: On evaporation from the surface of river basins, *Trans. Met. Obs. Iur-evskogo, Univ. Tartu* 4, 1911 (in Russian).
- Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, 99, 7866–7873, doi:10.1007/S00704-009-0134-9, 2010a.
- Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, *J. Hydrol.*, 395, 199–215, 2010b.
- Pike, J. G.: The estimation of annual runoff from meteorological data in a tropical climate, *J. Hydrol.*, 2, 116–123, 1964.
- Potter, N. J. and Zhang, L.: Interannual variability of catchment water balance in Australia, *J. Hydrol.*, 369, 120–129, 2009.
- Schmidli, J., Frei, C., and Vidale, P.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, *Int. J. Climatol.*, 26, 679–689, 2006.
- Schreiber, P.: Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa, *Meteorol. Z.*, 21, 441–452, 1904.
- Sankarasubramanian, A. and Vogel, R. M.: Annual hydroclimatology in the United States, *Water Resour. Res.*, 38, 1083, doi:10.1029/2001WR000619, 2002.
- Sankarasubramanian, A. and Vogel, R. M.: Hydroclimatology of the continental United States, *Geophys. Res. Lett.*, 30, 1363, doi:10.1029/2002GL015937, 2003.
- Sharma, D., Das Gupta, A., and Babel, M. S.: Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand, *Hydrol. Earth Syst. Sci.*, 11, 1373–1390, doi:10.5194/hess-11-1373-2007, 2007.
- Silberstein, R. P.: Hydrological models are so good, do we still need data?, *Environ. Modell. Softw.*, 21, 1340–1352, 2006.
- Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P.: The ability of a GCM-forced hydrological model to reproduce global discharge variability, *Hydrol. Earth Syst. Sci.*, 14, 1595–1621, doi:10.5194/hess-14-1595-2010, 2010.
- Sperna Weiland, F. C., van Beek, L. P. H., Weerts, A. H., and Bierkens, M. F. P.: Extracting information from ensemble of GCMs to reliably assess future global runoff change, *J. Hydrol.*, 412–413, 66–75, 2012.
- Strzepek, K. M. and Yates, D. N.: Climate change impacts on the hydrologic resources of Europe: a simplified continental scale analysis, *Climatic Change*, 36, 79–92, 1997.
- Terink, W., Hurkmans, R. T. W. L., Torfs, P. J. J. F., and Uijlenhoet, R.: Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin, *Hydrol. Earth Syst. Sci.*, 14, 687–703, doi:10.5194/hess-14-687-2010, 2010.
- Thornton, P. E., Running, S. W., and White, M. A.: Generating surfaces of daily meteorology variables over large regions of complex terrain, *J. Hydrol.*, 190, 214–251, 1997.
- Turc, L.: Le bilan d'eau des sols. Relation entre la précipitation, l'évaporation et l'écoulement, *Ann. Agron.* 5, 491–569, 1954.
- Varis, O., Kajander, T., and Lemmelä, R.: Climate and water: from climate models to water resources management and vice versa, *Climatic Change*, 66, 321–344, 2004.
- Vicuna, S. and Dracup, J. A.: The evolution of climate change impact studies on hydrology and water resources in California, *Climatic Change*, 82, 327–350, 2007.
- Vogel, R. M., Wilson, I., and Daly, C.: Regional regression models of annual streamflow for the United States, *J. Irrig. Drain. E.-ASCE*, 125, 148–157, 1999.
- Von Storch, H., Zorita E., and Cubasch, U.: Downscaling of global climate change to regional scales: an application to Iberian rainfall in wintertime, *J. Climate*, 6, 1161–1171, 1993.
- Wei, H., Gutowski, W. J., Vorosmarty, C. J., and Fekete, B. M.: Calibration and validation of a regional climate model for pan-Arctic hydrologic simulation, *J. Climate*, 15, 3222–3236, 2002.
- Wilby, R. L., Hay, L. E., Gutowski Jr., W. J., Arriitt, R. W., Takle, E. S., Pan, Z., Leavesley, G. H., and Clark, M. P.: Hydrological responses to dynamically and statistically downscaled climate model output, *Geophys. Res. Lett.*, 27, 1199–1202, 2000.
- Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., and Mearns, L. O.: Guidelines for use of climate scenarios developed from statistical downscaling methods, IPCC task group on data and scenario support for impacts and climate analysis, 2004.
- Willmott, C. J.: On the validation of models, *Phys. Geogr.*, 2, 184–194, 1981.
- Willmott, C. J.: Some comments on the evaluation of model performance, *B. Am. Meteorol. Soc.*, 63, 1309–1369, 1982.
- Willmott, C. J., Ackleson, S. G., Davis, R. E., Fiedema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., and Rowe, C. M.: Statistics for the Evaluation and Comparison of Models, *J. Geophys. Res.*, 90, 8995–9005, 1985.
- Wisser, D., Fekete, B. M., Vörösmarty, C. J., and Schumann, A. H.: Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H), *Hydrol. Earth Syst. Sci.*, 14, 1–24, doi:10.5194/hess-14-1-2010, 2010.
- Wolock, D. M. and McCabe, G. J.: Explaining spatial variability in mean annual runoff in the conterminous United States, *Clim. Res.*, 11, 149–159, 1999.
- Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long range experimental hydrologic forecasting for the eastern U.S., *J. Geophys. Res.*, 107, 4429, doi:10.1029/2001JD000659, 2002.
- Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrological implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, 62, 189–216, 2004.
- Yates, D. N.: Approaches to continental scale runoff for integrated assessment models, *J. Hydrol.*, 201, 289–310, 1997.
- Zhang, L., Daws, W. R., and Walker, G. R.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, *Water Resour. Res.*, 37, 701–708, 2001.